Alternative Approaches to Potential of Mean Force Calculations: Free Energy Perturbation versus Thermodynamic Integration. Case Study of Some Representative Nonpolar Interactions

CHRISTOPHE CHIPOT[*,†] **and PETER A. KOLLMAN**
Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

DAVID A. PEARLMAN
Vertex Pharmaceuticals Incorporated, 40 Allston Street, Cambridge, Massachusetts 02139-4211

Received 28 March 1995; accepted 6 September 1995

ABSTRACT

We investigated the convergence behavior of potential of mean force (PMF) calculations using free energy perturbation (FEP), thermodynamic integration (TI), and "slow growth" (SG) techniques. The critical comparison of these alternative approaches is illustrated by the study of three different systems: two tagged argon atoms in a periodic box of argon, two methane molecules, and two benzene molecules maintained in a "T-shaped" conformation, both dimers embedded in a periodic box of water. The complete PMF simulations were carried out considering several protocols, in which the number of intermediate "λ" states, together with the amount of sampling per individual state, were varied. In most cases, as much as 1 ns of molecular dynamics (MD) sampling was used to derive each free energy profile. For the different systems examined, we find that FEP and TI unquestionably constitute robust computational methods leading to results of comparable accuracy. We also show that proper convergence

*Author to whom all correspondence should be addressed at his present address: NASA-Ames Research Center, Moffett Field, CA 94035-1000.

†On leave from Laboratoire de Chimie Théorique, Unité de Recherche Associée au CNRS No. 510, Université Henri Poincaré–Nancy I, BP. 239, 54506 Vandœuvre-lès-Nancy, Cedex, France.

Journal of Computational Chemistry, Vol. 17, No. 9, 1112–1131 (1996)

CCC 0192-8651 / 96 / 091112-20

of the free energy calculations, and further quantitative interpretation of the PMFs, requires total simulation times much higher than has been hitherto estimated. In some circumstances, the free energy profiles derived from FEP calculations tend to be slightly poorer than those obtained with TI, as a probable consequence of the greater sensitivity of FEP to the window spacing $\delta\lambda$. In the context of TI, and to a lesser extent FEP, simulations, it appears preferable to employ a limited number of "λ" points of the integrand involving extensive sampling, rather than numerous points with fewer samplings. Finally, we note that, at least in the case of nonpolar interactions, PMFs of reasonable quality can be generated using SG, and at a substantially lower cost than with either FEP or TI.

Introduction

For the past 20 years, the computation of both free energy differences and absolute free energies on nontrivial systems have raised a great deal of attention.[1-13] In spite of their apparent success, these calculations cannot be considered as "black box" routine jobs. In particular, until very recently[14-22] little attention had been devoted to both the reliability and the accuracy of free energy simulations. Such calculations are based on the averaging of a quantity over an ensemble of states generated by either classical molecular dynamics (MD)[23] or Monte Carlo simulations (MC).[24] While several free energy computations reported in the literature employed 10–100 ps of sampling, it has been observed that transformations like solute → nothing in solution, commonly performed to determine a free energy of solvation, could require as much as 400–800 ps to yield an acceptable statistical accuracy.[14,22]

The choice of the method employed to estimate the changes in free energy may affect the efficiency of the calculation,[21] and, in turn, the amount of sampling necessary to reach proper convergence. Two of the most popular techniques are free energy perturbation (FEP)[25,26] and thermodynamic integration (TI).[26-28] In the first approach, the free energy difference between two states a and b is expressed rigorously by:

$$\Delta G = G_b - G_a = -RT \ln\langle e^{-[H_b(\mathbf{r}) - H_a(\mathbf{r})]/RT}\rangle_a \quad (1)$$

assuming that, in practice, state b is not too dissimilar from state a, so that the difference between the two states may be considered as a perturbation. $H_a(\mathbf{r})$ and $H_b(\mathbf{r})$ are the classical Hamiltonians corresponding to states a and b, respectively, depending on the atomic coordinate set $\{\mathbf{r}\}$ only. R is the molar gas constant and T is the absolute temperature. $\langle\ \rangle_a$ stands for the ensemble average reflecting state a.

In the case where states a and b are drastically different, the statistical ensemble generated using $H_a(\mathbf{r})$ will reflect only a small number of configurations where $H_b(\mathbf{r})$ is low enough in energy. This "orthogonality" of the Hamiltonians can be alleviated by considering a nonphysical pathway, linking state a to state b, by means of a reasonable number of intermediate states, or "windows,"[3,6] defined by a coupling constant λ. In doing so, the λ-dependent potential energy function varies monotonically from $H(\mathbf{r};\ \lambda = 0) = H_a(\mathbf{r})$ to $H(\mathbf{r};\ \lambda = 1) = H_b(\mathbf{r})$. If N windows are used over the entire simulation, the total free energy difference is evaluated from the sum of individual free energy differences between contiguous "λ" states:

$$\Delta G = \sum_{i=1}^{i=N} \Delta G_i = \sum_{\lambda=0}^{\lambda=1} -RT \ln\langle e^{-[H(\mathbf{r};\ \lambda+\delta\lambda) - H(\mathbf{r};\ \lambda)]/RT}\rangle_\lambda. \quad (2)$$

In the second approach to free energy calculations, the free energy difference between states a and b simply follows from the definition of an integral and the statistical connection between G and $H(\mathbf{r};\ \lambda)$[3]:

$$\Delta G = \int_0^1 \frac{\partial G(\lambda)}{\partial \lambda}\, d\lambda = \int_0^1 \left\langle \frac{\partial H(\mathbf{r};\ \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (3)$$

where $H(\mathbf{r};\ \lambda)$, λ, and $\langle\ \rangle_\lambda$ are defined as above.

While FEP has been employed extensively in the past few years, TI has recently become the subject of increasing interest[15,19,29,30] for essentially two major reasons: because of its formulation, the possible orthogonality of the Hamiltonians charac-

terizing states a and b is not an obstacle in the computation of the free energy; and unlike FEP, which uses the logarithm of the average of an exponential function, the free energy difference is evaluated from the ensemble average directly. In practice, the continuous integral (3) is approximated by a discrete sum, using, for instance, the trapezoidal rule or a Gauss–Legendre quadrature. This approximation constitutes the only drawback of the TI method, but is not critical as long as the integrand varies smoothly and is evaluated at a sufficiently large number of "λ" points.

Often referred to as a separate method, "slow growth" (SG) actually corresponds to the limiting case of FEP or TI, assuming that the window width $\delta\lambda$ is small enough to ensure a quasistatic evolution of the system. The ensemble average is reduced to a single value and the derivative (3) is approximated by a finite difference:

$$\left\langle \frac{\partial H(\mathbf{r};\lambda)}{\partial \lambda} \right\rangle d\lambda \simeq \frac{\partial H(\mathbf{r};\lambda)}{\partial \lambda} d\lambda = \Delta H. \quad (4)$$

Another important feature conditioning the success of free energy calculations is the choice of the nonbonded parameters for the molecular simulations. In particular, it has been recognized that the definition of the electrostatic model may play a key role in the description of intermolecular interactions between polar compounds.[22,31,32] Conversely, in the case of nonpolar species, where the Lennard–Jones contribution is the predominant term of the total interaction energy, the inappropriate calibration of the van der Waals parameters leads to free energies in poor agreement with the experiment.[22]

Frequently overlooked, the characterization of both the accuracy and the reliability of free energy calculations constitutes a topical problem. Recent studies have focused their attention on the convergence behavior of these simulations,[14–20] leading to the conclusion that reliably converged results can be obtained at the cost of intensive sampling. Until recently, the accuracy of most computations was estimated by means of three different approaches: (i) the similarity of the results derived from two distinct free energy simulations involving different amounts of sampling (e.g., 200 and 400 ps) gives a rough idea whether the convergence of the trajectories has been reached, (ii) the "hysteresis" between the simulation carried out from state a to state b, and that from state b to state a may also indicate that a proper convergence has been attained. (In practice, this method, as well as the related "double-wide" sampling technique,[33,34] provides a lower bound of the actual error bar. If, for instance, numerous "λ" points are used, involving a very limited number of equilibration and data collection steps, the difference between the "forward" and the "reverse" values of the free energy will frequently be artificially small, regardless of the actual quality of the computation.) (iii) The mean value and the root mean square deviation of the free energies computed using different starting configurations (i.e., generation of the positions and velocities from MD equilibrations of various lengths, or assignment of the initial velocities with different random number seeds) provide additional, and certainly more realistic, clues concerning the accuracy of the simulations.

It is now widely admitted that none of these approaches is totally satisfactory, essentially because they do not offer any clear suggestion on how to improve the efficiency of the simulations. In contrast, a new method based on the evaluation and the statistical analysis of free energy derivatives[20,35] allows the convergence behavior of the simulations to be probed reliably, giving pertinent information related to the simulation parameters used, and how these parameters should be modified to yield better quality results. Convergence of these derivatives can be quantified using statistical analysis of the series data, as has recently been described and applied to the ensemble data in free energy calculations.[35]

This article is the continuation of a study that was initiated[19] on a system consisting of two neon atoms in a periodic box of water. In the present contribution, we critically compare the FEP, TI, and SG methods for determining the potential of mean force (PMF) of two tagged argon atoms in a box of argon, two methane molecules in an aqueous medium, and two benzene molecules, arranged in a T-shaped motif, also embedded in a periodic box of water, putting the emphasis on the systematic examination of the convergence behavior of the simulations. The first system of interest was chosen, not only because of its intrinsic simplicity, but mainly because it allows the derivation of an "exact" PMF from the corresponding solvent–solvent radial distribution function (RDF), readily available from a conventional MD simulation. The prototypical hydrated methane dimer constitutes a straightforward, yet nontrivial application of FEP, TI, and SG for generating a PMF to an all-atom model. Apart from the fact that it probably represents one of the simplest "π–π"

interactions, the third molecular system was chosen to estimate the influence of including additional internal constraints, such as angular constraints to maintain the aromatic rings in an "edge-to-face" conformation, on the performance of the free energy calculations. After summarizing in the following section the technical details of the different molecular simulations, we discuss the merits of each computational method on the basis of a thorough analysis of its inherent convergence properties.

Methodology: Computational Details

To assess the validity of the FEP, TI, and SG approaches in the framework of PMF simulations, we first considered a system of two tagged argon atoms in liquid argon. The model used for this purpose consisted of a periodic box containing 98 argon atoms (approximately $21.8 \times 15.5 \times 15.5$ Å^3). The Lennard–Jones parameters employed here were taken from the earlier study of Tobias and Brooks[36] (see Table I), leading to a density number of $\rho \simeq 0.020$ Å^{-3}. On account of its relative simplicity, this system allows long simulations (typically greater than 1 ns) to be carried out at a very reasonable cost. In principle, if infinite MD sampling is performed, the PMF $w(\mathbf{r})$, derived from an FEP, TI, or SG simulation, is expected to converge exactly toward the logarithm of the RDF $g(\mathbf{r})$:

$$w(\mathbf{r}) = -k_B T \ln g(\mathbf{r}). \tag{5}$$

TABLE I.
Nonbonded Parameters Used in Molecular Simulations.

	Atom Type	Charges (ecu)	Lennard–Jones Parameters	
			R_{ii}^* (Å)	ε_{ii} (kcal / mol)
	Ar[a]	0.000	1.9110	0.2381
H_2O[b]	OW	−0.834	1.7680	0.1520
	HW	0.417	0.0000	0.0000
CH_4[c]	CT	−0.560	1.8000	0.0600
	HC	0.140	1.3750	0.0380
C_6H_6[d]	C	−0.138	1.9080	0.0860
	HA	0.138	1.4590	0.0150

[a] Lennard–Jones parameters of argon by Tobias and Brooks.[36]
[b] TIP3P water model by Jorgensen et al.[37]
[c] Original AMBER all-atom force field.[39,43]
[d] Second generation AMBER all-atom force field.[41,42]

The implication of the statistical relation between $w(\mathbf{r})$ and $g(\mathbf{r})$ is particularly important in the sense that it constitutes an unambiguous criterion to probe the convergence of our free energy calculations. In practice, since infinite sampling is utterly unrealistic, we attempted to estimate, through different simulation strategies, the optimum protocol capable of offering the best possible reproduction of the assumed exact PMF. In other words, our main goal is to obtain a free energy profile that can be superimposed to the logarithm of the RDF, without any vertical adjustment to make the minima coincide.[36]

For the second application of FEP, TI, and SG to the derivation of free energy profiles, we considered a system of two methane molecules embedded in a periodic rectangular box of 394 explicit TIP3P[37] water molecules (approximately $21.5 \times 21.5 \times 30.0$ Å^3), whereas for the third one, two benzene molecules were immersed in a periodic box containing 568 TIP3P water molecules (approximately $31.7 \times 23.4 \times 24.0$ Å^3).

All the PMF computations were carried out using the molecular simulation package GIBBS/AMBER 4.1.[38] The potential energy function employed in the course of the various simulations is of the form[39–42]:

$$\begin{aligned} H_{\text{total}}(\mathbf{r}) &= \sum_{\text{bonds}} k_r (r - r_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 \\ &+ \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \omega)] \\ &+ \sum_{i<j} \left\{ \varepsilon_{ij} \left[\left(\frac{R_{ij}^*}{R_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^*}{R_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon R_{ij}} \right\} \end{aligned} \tag{6}$$

and the mixing rules are $\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$ and $R_{ij}^* = \frac{1}{2}(R_{ii}^* + R_{jj}^*)$.

Because they have led to reasonable agreement with experiment in hydration free energy calculations,[22] the Weiner et al.[39,40] and Hagler et al.[43] van der Waals parameters, supplemented by potential derived net atomic charges[44,45] obtained from a 6-31G**[46] wave function, were employed to model the methane dimer (see Table I). Conversely, in the case of the "T-shaped" benzene dimer, we adopted the Lennard–Jones parameters recently developed by Cornell et al.,[41,42] to which 6-31G** potential derived charges were added (see Table I).

All the MD simulations presented herein were carried out in the isobaric–isothermal (n, P, T) ensemble. The time step for integrating the trajectories was set to 1.0 or 2.0 fs, depending on the simulation; and the average temperature and pressure were maintained, in the case of the hydrated methane and benzene dimers, at 300 K and 1 atm, respectively, employing the Berendsen algorithm[47] with a separate temperature coupling for the solutes and the solvent. Unfortunately, this algorithm proved to be extremely slowly convergent for systems consisting of weakly interacting particles. Therefore, we maintained the average temperature of the argon periodic box at 86 K, by means of a stochastic reassignment of the velocities.[48] Long-range interactions were truncated using a hard cutoff of 8.0 Å, in the case of the argon–argon PMF, and of 9.0 Å, for the methane–methane and benzene–benzene PMFs.

All the chemical bonds were constrained to their equilibrium value by means of the SHAKE[49,50] procedure. Prior to the various PMF simulations, the different systems of interest were equilibrated thoroughly: 200, 25, and 25 ps, for the argon dimer, methane dimer, and benzene dimer, respectively.

At every "λ" point, the distance separating the two solutes was kept fixed using a distance holonomic constraint.[51] The fixed interatomic distance was changed continuously from 7.7 to 3.0 Å, 8.0 to 2.5 Å, and 9.5 to 4.0 Å, in the case of the argon dimer, the methane dimer, and the benzene dimer, respectively, hence constituting the overall reaction path of the different PMFs. In the first system, the SHAKE-constrained intersolute distance corresponds to the separation of the two tagged argon atoms. For the methane–methane complex, a similar internal constraint is applied between the two carbon atoms (see Fig. 1). The situation of the T-shaped benzene dimer is slightly more complex, since, in addition to the fixed distance separating the noninteracting centroids of the two aromatic rings (see Fig. 1), a set of appropriately chosen angular holonomic constraints was incorporated to maintain the solutes perpendicular to each other along the entire reaction path; i.e., six constraints $\{C_i, D_1, D_2\}$ were set to 90.0°, together with three constraints $\{D_1, D_2, C_j\}$ set to 60.0 and 120.0° (see Fig. 1). Depending on the computational method employed to generate the PMF profiles, the number of windows and the amount of sampling per window were varied.

In the case of the FEP methodology, the putative critical influence of the number of nonphysical "λ" states on the overall aspect of the generated

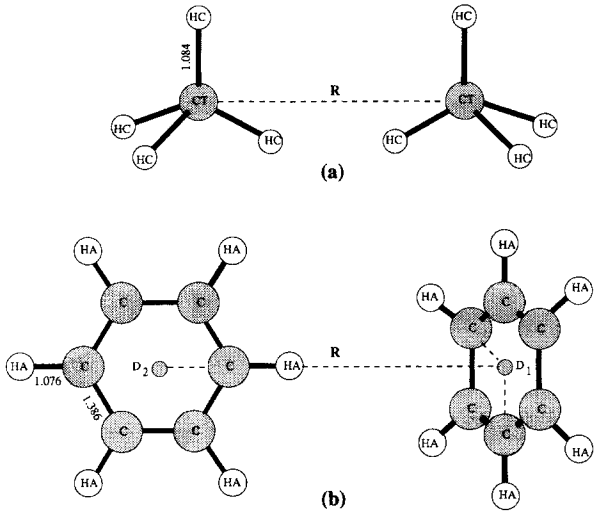

FIGURE 1. Geometrical parameters used for the molecular simulations of (a) the hydrated methane dimer, and (b) the hydrated "T-shaped" benzene dimer. All bond lengths in Å. In the case of the benzene dimer, two pseudo noninteracting atoms D_1 and D_2, located at the center of the aromatic rings, are used to define both the constrained distance R, and the constrained angles $\{C_i, D_1, D_2\}$ and $\{D_1, D_2, C_j\}$, involved in the PMF calculations.

PMFs[19] led us to consider a series of different sampling protocols. In particular, for each system, we examined the cases corresponding to a modest number of windows with extensive MD sampling per intermediate state, and, reciprocally, a large number of windows involving fewer equilibration and data collection steps. The different sampling strategies employed for the three systems of interest are given in Table II. Since the constrained interatomic distance between the two solutes is a λ-dependent variable of the potential energy function $H(\mathbf{r})$, the so-called "PMF bond contribution"[14] was evaluated along the reaction path, i.e., for each configuration corresponding to the potential energy $H_a(\mathbf{r}) = H[\mathbf{r}; \lambda; r_{ij}^{\text{constraint}}(\lambda)]$, $H_b(\mathbf{r}) = H[\mathbf{r}; \lambda + \delta\lambda; r_{ij}^{\text{constraint}}(\lambda + \delta\lambda)]$ is evaluated by modifying the constrained distance to its appropriate value at $\lambda + \delta\lambda$. It should be noted that, prior to the computation of the potential energy for this state, the atoms affected by translation along the constrained bond are moved.

Within the context of the TI approach, the sampling protocols adopted were mostly identical to those selected for the FEP simulations (see Table II). The λ dependence of the constrained distance separating the solutes is introduced in the TI formulation by means of the potential force (PF)

TABLE II.
Sampling Protocols Used in Free Energy Simulations.

Simulation	Method	$\delta\lambda$	Sampling (ps)		Total Sampling (ps)
			Equilibration	Collection	
Ar—Ar					
1a	FEP	0.04	14.0	14.0	700.0
1b	TI	0.04	14.0	14.0	700.0
2a	FEP	0.02	30.0	30.0	3000.0
2b	TI	0.02	30.0	30.0	3000.0
3a	FEP	0.004	30.0	30.0	15000.0
3b	TI	0.004	30.0	30.0	15000.0
4a	FEP	0.001	0.5	0.5	1000.0
4b	TI	0.001	0.5	0.5	1000.0
5	SGTI	0.00001	0.0	0.004	400.0
6	SGTI	0.000001	0.0	0.004	4000.0
CH_4—CH_4					
7a	FEP	0.05	1.0	1.5	50.0
7b	TI	0.05	1.0	1.5	50.0
8a	FEP	0.05	4.0	6.0	200.0
8b	TI	0.05	4.0	6.0	200.0
9a	FEP	0.01	0.5	0.5	100.0
9b	TI	0.01	0.5	0.5	100.0
10a	FEP	0.05	10.0	15.0	1000.0
10b	TI	0.05	10.0	15.0	1000.0
11a	FEP	0.02	5.0	15.0	1000.0
11b	TI	0.02	5.0	15.0	1000.0
12a	FEP	0.005	2.0	3.0	1000.0
12b	TI	0.005	2.0	3.0	1000.0
13a	FEP	0.001	0.5	0.5	1000.0
13b	TI	0.001	0.5	0.5	1000.0
13c	FEP[a]	0.001	0.5	0.5	1000.0
14	FEP	0.0002	0.25	0.25	2500.0
15	SGTI	0.00001	0.0	0.002	200.0
C_6H_6—C_6H_6					
16a	FEP	0.01	2.5	7.5	1000.0
16b	TI	0.01	2.5	7.5	1000.0
17a	FEP	0.01	1.25	3.75	500.0
17b	TI	0.01	1.25	3.75	500.0
18a	FEP	0.05	10.0	40.0	1000.0
18b	TI	0.05	10.0	40.0	1000.0
19a	FEP	0.05	10.0	15.0	500.0
19b	TI	0.05	10.0	15.0	500.0
20	SGTI	0.00001	0.0	0.002	200.0

[a] Identical simulation strategy, but using a different random number seed.

method,[19] which allows the determination of $\partial H^{\text{constraint}}/\partial\lambda$, the derivative due to the holonomic constraint $r_{ij}^{\text{constraint}}$. This method utilizes the most natural definition for the generation of PMFs, that is, the average of the Cartesian forces, which are evaluated at each MD step. Unlike the PMF bond contribution method introduced in FEP calculations, the TI/PF approach does not require two evaluations at each integration point of the nonbonded interactions for those atoms, the positions of which are dependent on the constraints. Regarding the computational cost of the PMF simulation, this feature appears to be particularly important.

Finally, in the case of the SG procedure, we employed the TI/PF methodology, using 100,000

or 1,000,000 "λ" points, depending on the case of the integrand, and generally limiting the sampling to one integration point of data collection per window. Since this approach assumes that the system evolves quasistatically (i.e., the magnitude of the perturbation within the box between states λ and $\lambda + \delta\lambda$ is negligible), no equilibration is *a priori* required. In the following, this third computational method will be referred to as SGTI.

Results and Discussion

ARGON–ARGON PMF

The computed free energy profiles characterizing the mutual approach of two tagged argon atoms in liquid argon are shown in Figure 2a. These PMFs, derived from simulations 3a, 3b and 6 (see Table II), corresponding, respectively, to the three alternative approaches FEP, TI, and SGTI, are very similar and almost coincide with the solvent–solvent RDF, in fact, $-k_BT \ln g(\mathbf{r})$, of this system. This result is not completely surprising in view of the amount of MD sampling involved in each calculation. Indeed, while the SGTI simulation was carried out over 4 ns, both FEP and TI PMFs required as much as 15 ns individually. Perhaps the curve resulting from the SGTI is the least satisfactory, in light of its slightly diverging tail end, around 7.0 Å. It is, nevertheless, worth noting that, at variance with the free energy profile reported by Tobias and Brooks,[36] both the "contact" and the "solvent-separated" minimum occur at the correct interatomic distances (approximately 3.7 and 7.2 Å, respectively), with the appropriate well depths (approximately -0.18 and -0.04 kcal/mol, respectively). The pertinent test that consists of comparing $g(\mathbf{r})$ and $w(\mathbf{r})$, in order to analyze the convergence behavior of the free energy simulations, confirms the integrity of both the PMF bond contribution and TI/PF methods for calculating the free energy component arising from internal constraints, in the context of PMF calculations.

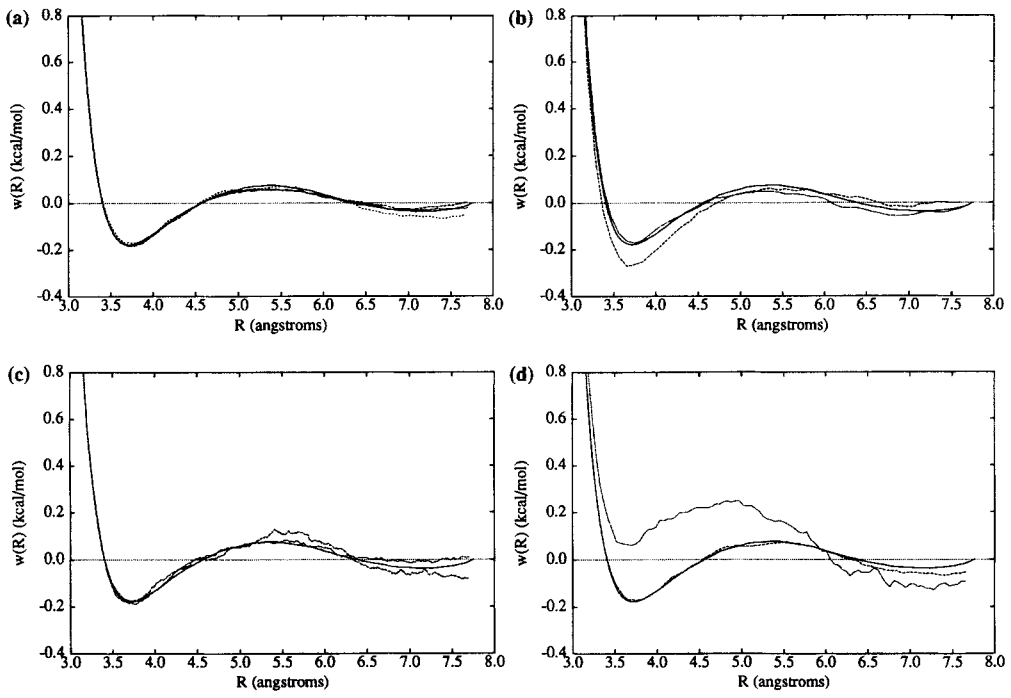

FIGURE 2. Argon–Argon PMF evaluated using: (a) FEP (– – –), TI (· · ·), over 250 windows of 30-ps equilibration +30-ps data collection, and SGTI (---) over 1,000,000 windows of 4-fs data collection; (b) FEP (---) and TI (· · ·) over 50 windows of 30-ps equilibration + 30-ps data collection; (c) FEP (---) and TI (· · ·) over 1000 windows of 0.5-ps equilibration + 0.5-ps data collection; (d) SGTI over 1,000,000 (---) and 100,000 (· · ·) windows of 4-fs data collection. All the profiles are generated as a function of R, the distance separating the two tagged argon atoms, and are compared with the exact PMF (—) derived from a solvent–solvent RDF.

At this stage, we observed that FEP and TI, and to a certain extent SGTI, yield identical results *if* extensive sampling was performed. What can be expected from a noticeable decrease of the number of intermediate "λ" states? The PMFs shown in Figure 2b were generated over 50 windows, using 30 ps of equilibration followed by 30 ps of data collection per window (i.e., simulations 2a and 2b, described in Table II). It can now be witnessed that while the overall shape of the profile obtained with TI is still in reasonable accord with the "exact" PMF, the FEP curve markedly departs from it. In particular, the depth of the "close-contact" minimum appears to be overestimated by ca. 0.1 kcal/mol. This result is somewhat troubling if one considers that the total length of the simulation is equal to 3 ns. Moreover, the extensive equilibration at each "λ" state clearly suggests that a lag between the Hamiltonians representative of contiguous states is unlikely to be the reason of the discrepancy. Whether the lack of agreement between the FEP free energy profile and the logarithm of the RDF could be ascribed to the possible orthogonality of the Hamiltonians is not clear yet.

In contrast with the preceding simulations, we substantially increased the number of windows over the entire run, while reducing the amount of sampling per window accordingly (cf. simulations 4a and 4b). As may be seen in Figure 2c, the curves generated with both the TI and the FEP approaches roughly show similar characteristics. The FEP profile, however, is significantly noisier than the TI one. In fact, while the "contact" minimum in the FEP curve is reasonably reproduced, the free energy barrier occurring around 5.5 Å is slightly overestimated, and it is virtually impossible to discern a "solvent-separated" minimum beyond 6.5 Å.

An additional series of simulations were carried out, using a very limited number of intermediate "λ" states: i.e., 25 windows of 14 ps of equilibration followed by 14 ps of production dynamics (cf. simulations 1a and 1b). Compared to the targeted exact curve, both the FEP and TI free energy profiles (data not shown) are inaccurate, yet retain key qualitative features. The intrinsic noise of the PMFs, in conjunction with the limited number of windows, makes the assignment of a "solvent-separated" minimum difficult. We note that the TI curve, although less noisy than the FEP one, tends to diverge more rapidly beyond 6.0 Å.

Finally, in the context of SGTI simulations, we analyzed two strategies, employing, respectively, 100,000 and 1,000,000 windows (i.e., simulations 5 and 6, described in Table II). As may be observed in Figure 2d, only the first sampling protocol yields a satisfactory result. In contrast, the conspicuously repulsive free energy profile derived from simulation 5 is totally unacceptable, albeit presenting reminiscences of both the "contact" and the "solvent-separated" minima. Clearly, these two curves illustrate the problems inherent to the lag between the Hamiltonians characterizing contiguous "λ" states. It is probable that in simulation 5, either the number of windows or the amount of sampling (in the light of the aforementioned results) is insufficient. In fact, what seems to be happening here is a propagation, and to a certain extent an amplification, of the error along the reaction path, leading to a poor reproduction of the exact PMF.

So far, we have seen that in the perspective of determining PMF profiles, FEP and TI give roughly identical results under the *sine qua non* condition that enough MD sampling has been performed. Furthermore, the curves presented in this section provide a potent example of the necessity to perform adequate sampling to attain properly converged results, as has been highlighted earlier.[19] If this is the case (see Fig. 2a), FEP and TI are indeed unquestionably equivalent. If not, it would seem that, in most circumstances, TI appears to be slightly superior to FEP. In our opinion, this is likely to be rooted in the orthogonality of adjacent states, especially when a limited number of windows is employed over the entire simulation. The case of SGTI is particularly instructive in the sense that if the amount of MD sampling is limited to 4 fs per window, as many as 1 million windows are necessary to yield seemingly converged results, representing a variation of the constrained interatomic distance slightly smaller than $5\ 10^{-6}$ Å.

What emerges from this preliminary study is that *only* very long simulations (of total length typically greater than 1 ns) can ensure accurate PMF. Considering the trivial nature of the present system, which limits the cost of the calculations (see Table III), this fact is somewhat preoccupying, and tends to suggest that even longer simulations, hence higher CPU investments, might be required to yield acceptable free energy profiles for more complex systems. This may not necessarily always be true. In the particular case of liquid argon, the reduced kinetic energy of the atoms at 86 K, compared to liquid water at 300 K, slows down the relaxation process between contiguous "λ" states of the reaction path. Therefore, it is not completely unexpected that the amount of MD sampling per

TABLE III.
Typical CPU Times Involved in Free Energy Simulations.

Simulation	Method	Total Sampling (ps)	Contributions[a] Constraints (%)	Contributions[a] PMF (%)	CPU Time (h)
Ar—Ar					
2a	FEP	3000.0	0.7	9.5	10.78[b]
2b	TI	3000.0	0.6	0.5	9.24[b]
CH_4—CH_4					
7a	FEP	50.0	1.9	6.8	1.24[c]
7b	TI	50.0	2.0	0.1	1.16[c]
8a	FEP	200.0	—	—	2.03[d]
8b	TI	200.0	—	—	1.67[d]
9a	FEP	100.0	—	—	2.41[c]
9b	TI	100.0	—	—	2.24[c]
10a	FEP	1000.0	—	—	10.12[d]
10b	TI	1000.0	—	—	8.35[d]
15	SGTI	200.0	0.7	0.0	33.52[b]
C_6H_6—C_6H_6					
19a	FEP	1000.0	42.2	6.8	81.80[e]
19b	TI	1000.0	45.1	0.1	76.39[e]

[a]Percentage of the total CPU time spent in the dedicated subroutines.
[b]Measured on IBM RS-6000 / 375.
[c]Measured on SGI Challenge, single processor.
[d]Measured on SGI Challenge, four processors.
[e]Measured on SGI Challenge, two processors.

window required to generate a smooth PMF is substantially larger for liquid argon than for liquid water, as may be seen in the following section.

METHANE–METHANE PMF

In Figure 3a the methane–methane PMF are presented for the three alternative approaches FEP, TI, and SGTI (cf. simulations 11a, 11b, and 15). As noted above, while the two first simulations correspond to a total MD sampling time equal to 1 ns, the third one is reduced to 200 ps. From the onset, the overall smooth appearance of the different PMFs (in good agreement with previous observations on similar molecular systems[19,36,52–56]) tends to prove that the three methods give reasonably similar results. Qualitatively, the three generated curves show the characteristics of most free energy profiles derived for small hydrophobic molecules embedded in water: a global minimum at the "contact" distance, and a secondary minimum, often referred to as "solvent-separated," at a greater distance. Following the general trend emerging from original studies by Pangali and Berne[52] and Jorgensen et al.,[55] we note that, regardless of the computational approach, the "contact" dimer is more probable than the "solvent-separated" pair. This is also what Pratt and Chandler[57–59] have predicted on the basis of their theory for a system constituted of two hard spheres in water. At variance with our results, the MC simulation carried out by Ravishanker et al.[53] predicts the "solvent-separated" structure to be more favorable. As underlined by Jorgensen et al.,[55] this discrepancy is likely to be imputable to the use of a more attractive methane–water interaction potential.

A closer look at Figure 3a reveals that, while the free energy profiles generated using FEP and SGTI are reasonably similar, the one obtained from TI appears to be noticeably more attractive. In particular, the "contact" minimum of both the FEP and the TI PMF curves occurs at ca. 3.7 Å, with the free energies of −0.64 and −0.95 kcal/mol, respectively. On the other hand, the minima corresponding to the "solvent-separated" pair emerge at ca. 7.3 Å for FEP, and at 6.8 Å for TI, with the free energies of −0.17 and −0.08 kcal/mol, respectively. Because the present PMFs are determined from relative free energies, their absolute positioning should, in principle, fulfill the condition $w(\mathbf{r}) = 0$ at infinite separation of the solutes.[36] Because of the nature of these simulations, it is, however,

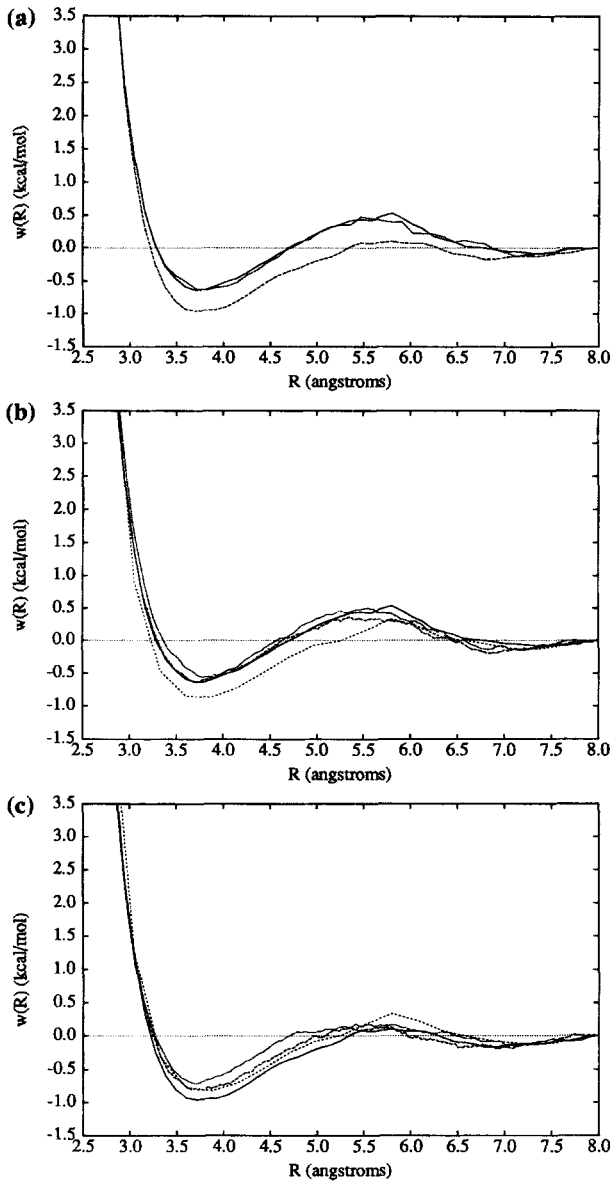

FIGURE 3. Methane–methane PMF evaluated using: (a) FEP (—), TI (---), over 50 windows of 5-ps equilibration +15-ps data collection, and SGTI (···) over 100,000 windows of 2-fs data collection; (b) FEP and (c) TI over 50 windows of 5-ps equilibration + 15-ps data collection (—), 20 windows of 10-ps equilibration + 15-ps data collection (---), 200 windows of 2-ps equilibration + 3-ps data collection (···), 1000 windows of 0.5-ps equilibration + 0.5-ps data collection (–––). All the profiles are generated as a function of R, the distance separating the carbon atoms.

legitimate to assume that 8.0 Å constitutes a reasonable separation for anchoring the PMF to zero. At this stage, it is not clear whether the ca. 0.3 kcal/mol free energy difference between the "contact" minima obtained via FEP and TI could be ascribed to an improper positioning of the curve or to a discrepancy in the convergence behavior of the two methods. In view of the -0.42 ± 0.34 kcal/mol free energy of binding reported by Jorgensen et al.,[55] for a system of two methane molecules separated by 4.0 Å in TIP4P water,[37] it would seem that our TI and, to a lesser extent, FEP and SGTI simulations have a tendency to lead to slightly exaggerated attractions. An evaluation of the free energy of binding for the present molecular system reveals, however, that this may not be necessarily true. Mutating $[(CH_4)_2]_{aqueous}$ into $[CH_4]_{aqueous}$ over 800 ps, using the electrostatic decoupling procedure[2] (400 windows per leg, involving 0.5 ps of equilibration and 0.5 ps of production dynamics at each window) and maintaining the two solutes at the 3.7 Å "contact" distance, leads to a compositional change in free energy of -1.45 ± 0.03 kcal/mol. On the other hand, a perturbation of $[CH_4]_{aqueous}$ into "nothing," carried out over 600 ps,[22] yields a free energy change equal to -2.26 ± 0.04 kcal/mol, in excellent agreement with the experiment.[60] We note in passing that these two mutations were performed using FEP, in conjunction with the PMF bond contribution[14] method to evaluate the component of the free energy due to shrinking bond lengths. The resulting net binding free energy of -0.81 ± 0.07 kcal/mol at "contact" separation—almost 0.4 kcal/mol lower than that reported by Jorgensen et al.[55] employing an identical approach—comfortably falls between the values provided by FEP and TI. It should be underlined here that the PMF presented by Jorgensen et al. was vertically adjusted by over 0.5 kcal/mol in order to match the correct binding free energy estimated at the contact intersolute distance.

In comparison with the results of Pangali and Berne,[52] Jorgensen et al.,[55] and Pratt and Chandler,[57] our "contact" minima emerge at slightly shorter intersolute separations. Moreover, the carbon–carbon distance found in this study for the "solvent-separated" minima bracket the values obtained by both Jorgensen et al. and Pangali and Berne. An estimation of the so-called "forward" free energy barrier, occurring at approximately 5.8 Å, leads to ca. 1.0 kcal/mol, in excellent accord with the value reported by Jorgensen et al., but somewhat greater than the ca. 0.56 kcal/mol of Pangali and Berne.

So far, it would seem that for equivalent sampling protocols, FEP and TI lead to analogous

results. That this may not be always the case, as can now be seen in Figure 3b and c. The different PMFs depicted in this figure were generated using the two alternative computational approaches FEP and TI, in association with the sampling strategies described in Table II. At first glance, the free energy profiles derived from FEP simulations 10a, 11a, 12a, and 13a, show similar characteristics (see Figure 3b). In general, the displayed curves are relatively smooth, possessing a "contact" minimum around 3.7 Å, and a "solvent-separated" one around 7.2 Å. It should be noted, however, that the position of these free energy minima fluctuates, depending on the simulation protocol adopted. In the case of the contact pair, the variation does not exceed ca. 0.2 Å; but as much as 0.5 Å can be witnessed for the "solvent-separated" minimum. Interestingly enough, the "contact" minimum obtained from simulation 10a (i.e., 20 windows of 10 ps equilibration, followed by 15 ps of data collection) is roughly 0.3 kcal/mol lower in free energy than those estimated using differing simulation protocols. It is not completely clear whether this fact could be related to the very limited number of windows involved in the calculation, and, hence, to the possible orthogonality of the Hamiltonians representative of contiguous "λ" states.

The set of curves reproduced in Figure 3c, derived from TI simulations 10b, 11b, 12b, and 13b, are also relatively smooth, showing analogous features. It can be observed that the most attractive PMF is obtained with 50 windows—5 ps of equilibration + 15 ps of data collection—(−0.95 kcal/mol deep contact minimum), while the least attractive one corresponds to 200 windows—2 ps of equilibration + 3 ps of data collection—(−0.69 kcal/mol). In each case, the "contact" minimum occurs around 3.7 Å, and the barrier separating the two minima emerges at ca. 5.8 Å. There is a similar agreement regarding the position of the "solvent-separated" pair—occurring around 6.8 Å with a corresponding free energy of −0.17 kcal/mol—although the corresponding portion of the curve generated with 1000 windows (0.5 ps of equilibration + 0.5 ps of data collection) appears to be somewhat noisier and less reliable. This remark may be correlated to the recent work of Pearlman[19] who suggested that it is generally preferable to employ a limited number of "λ" points involving extensive sampling rather than numerous points with fewer samplings. It is worth noting that in view of the free energy profile obtained with only 20 windows, TI seems to be less susceptible to suffer from a possible orthogonality between $H(\mathbf{r}; \lambda)$ and $H(\mathbf{r}; \lambda + \delta\lambda)$, albeit additional studies might be required to further support this statement.

At this stage, it appears from the various examples reported above, that both FEP and TI lead to a reasonable *consensus*. What really emerges from Figure 3b and c is that simulations of equal total length (1 ns), but involving different number of "λ" states and MD sampling per state, could lead to noticeably different PMF profiles. Since this effect turns out to be slightly more pronounced in the case of FEP, we considered a simulation protocol in which the number of windows was increased up to 5000, reducing the sampling to 0.25-ps equilibration and 0.25-ps data collection per window (cf. simulation 14). As may be observed in Figure 4a, the curve derived from this free energy calculation is very smooth, possessing both a "contact" and a "solvent-separated" minimum, occurring, respectively, at 3.7 and 7.0 Å, with the corresponding free energies of ca. −0.52 and −0.10 kcal/mol. It should be emphasized that this PMF is much smoother than the one obtained over 1000 windows, hence suggesting that FEP is particularly sensitive to the number of intermediate "λ" states used, but also calling into question the actual nature of the approach used: true FEP or pseudo SG/FEP? Moreover, a comparison of the free energy profiles derived from simulations 13a and 14—corresponding to 1 and 2.5 ns, respectively—tends to indicate that the latter is more likely to have converged than the first. In fact, repeating the PMF calculation using the sampling strategy of simulation 13a, but introducing a different random number seed (cf. simulation 13c), yields a markedly different curve, as may be witnessed in Figure 4a. While this profile clearly possesses a "contact" minimum located at 3.9 Å, corresponding to a free energy of −0.89 kcal/mol, the "solvent-separated" minimum is less distinct. In addition, the barrier separating the two minima appears to be much weaker than for any of the PMFs reported until now.

We have demonstrated that, even for seemingly long enough free energy simulations, the role of the sampling protocol, i.e., the optimal choice of both $\delta\lambda$ and the number of equilibration and data collection steps, may play a nonnegligible role on the general aspect of PMF. How does this translate for shorter simulation times, and which computational approach, FEP or TI, is more appropriate in this context? In an attempt to address these issues, we considered three different strategies, corresponding to 50–200 ps total simulation lengths. As

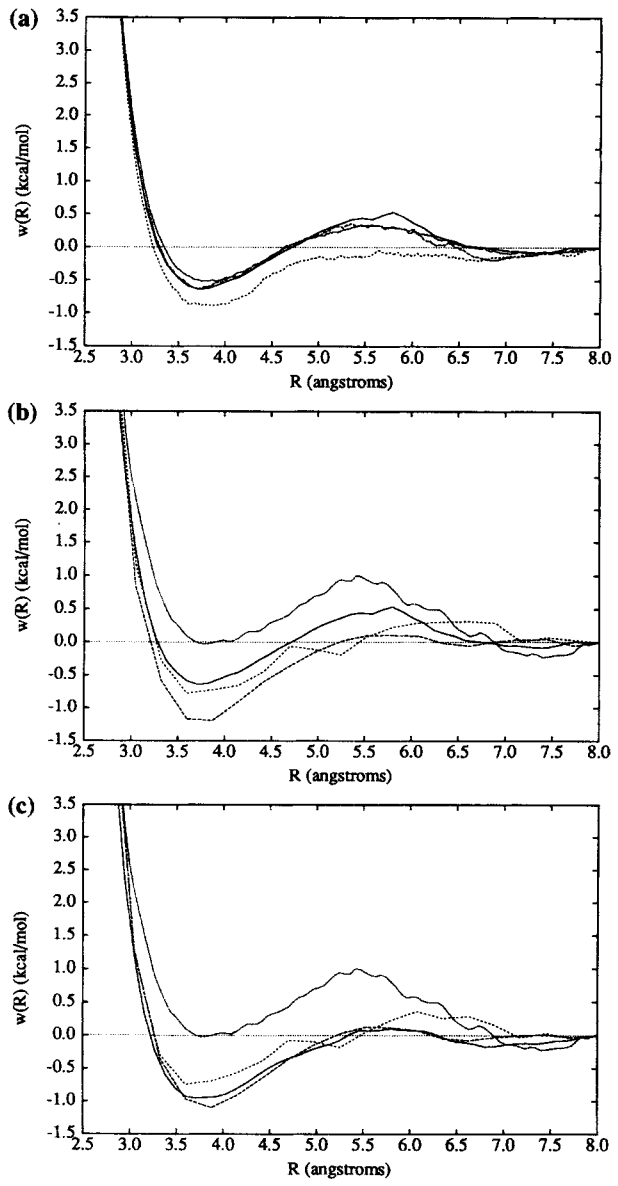

FIGURE 4. Methane–methane PMF evaluated using: (a) FEP, over 50 windows of 5-ps equilibration + 15-ps data collection (—), 1000 windows of 0.5-ps equilibration +0.5-ps data collection (– – –), identical sampling protocol, but introducing a different random number seed (---), 5000 windows of 0.25-ps equilibration +0.25-ps data collection (···); (b) FEP and (c) TI over 50 windows of 5-ps equilibration + 15-ps data collection (—), 20 windows of 1-ps equilibration + 1.5-ps data collection (---), 20 windows of 4-ps equilibration + 6-ps data collection (– – –), 1000 windows of 0.5-ps equilibration + 0.5-ps data collection (···). All the profiles are generated as a function of R, the distance separating the carbon atoms.

can be observed in Figure 4b, none of the PMFs—derived from FEP free energy simulations 7a, 8a, and 9a—tends to coincide with the curve obtained over 50 windows with more sampling (simulation 11a). On the contrary, the three profiles are unacceptably noisy, except perhaps for the one corresponding to the longest simulation time, i.e., 200 ps. The depth of its "contact" minimum is, nevertheless, overestimated by ca. 0.5 kcal/mol when compared to the curve characterizing simulation 11a. The two other free energy profiles, resulting from shorter simulations, constitute an excellent illustration of the lag between the Hamiltonians representative of adjacent "λ" states, and the propagation of the error along the reaction path. Clearly, both protocols for simulations 7a and 9a correspond to too small a number of windows for the amount of MD sampling performed.

In Figure 4c we have reproduced the free energy profiles obtained using TI, in association with similar sampling strategies (simulations 7b, 8b, and 9b). It is worth pointing out that, in the case of the two shortest calculations, the resulting PMFs are extremely poor and surprisingly almost identical to those derived employing FEP. On the other hand, the profile obtained from the 200 ps simulation is in much better accord with the reference simulation carried out over 50 windows and 1 ns (simulation 11b) than is its FEP homologue with simulation 11a.

In the light of the different PMFs presented in this section, it appears that FEP and TI can give relatively accurate results, of comparable quality, under the condition that appropriate sampling be performed. Conversely, in the case of limited MD sampling, both methods inevitably lead to mediocre free energy profiles, as a probable consequence of the lag between Hamiltonians $H(\mathbf{r}; \lambda)$ and $H(\mathbf{r}; \lambda + \delta\lambda)$. It should be mentioned, however, that FEP seems to be somewhat more sensitive than TI to the window spacing—hence, to the possible disparity of adjacent intermediate "λ" states (i.e., the relatively large magnitude of the changes being effected)—in conjunction with an inappropriate sampling. Regarding the SGTI approach, the results obtained herein are particularly promising, albeit at this stage, it has not been proven unequivocally that the SG is correct for the generation of PMF curves. In fact, the use of a sufficiently small window width $\delta\lambda$, and the underlying assumption that the system remains continually in an equilibrium state, should be considered carefully. It has been noted that, in the framework of SG simulations, the molecular system will

tend to "lag" the classical Hamiltonian $H(\mathbf{r}; \lambda)$.[61] Nevertheless, we note that for the hydrated methane dimer, the total CPU time for the SGTI calculation was ca. 5 times smaller than that involved with the standard TI procedure (due to the fact that only 1/5 as many MD steps were calculated, i.e., 200 ps). This makes SGTI a very appealing alternative to the CPU-intensive TI/FEP methods (see Table III), but it is not clear yet whether this approach will hold for solutes of greater complexity and/or polarity.

As already mentioned, probing the convergence of free energy simulations is a cumbersome task. In particular, in the absence of irrefutable tests, such as the logarithm of a converged RDF (*vide supra*), it is relatively difficult—not to say impossible—to give an accurate estimate of the error associated with the derived PMF profile. The nature of the TI formulation suggests an alternative approach to check whether the convergence of the simulation has been reached. Since the potential energy $H(\mathbf{r})$ is a function of a series of parameters $\{Q_i(\lambda)\}$ (nonbonded parameters R^*_{ij}, ε_{ij}, and q_i, or internal constraints), the first derivatives of the free energy with respect to any parameter $Q_i(\lambda)$ is expressed by[19,20]:

$$\frac{\partial G}{\partial Q_i(\lambda)} = \left\langle \frac{\partial H(\mathbf{r}; \lambda)}{\partial Q_i(\lambda)} \right\rangle. \tag{7}$$

Although the individual parameters are virtually unchanged along the pathway of the PMF simulation, it is possible to imagine infinitesimal mutations of the $\{Q_i(\lambda)\}$, suggestive of modifications that could be made to improve the free energy. More importantly for this work, monitoring these derivatives can provide a sensitive test of the rate of convergence of the calculated free energy. We evaluated the first derivatives of the free energy with respect to both the nonbonded parameters and the intersolute constraint distance along the reaction path. Apart from the time step that was set to 2.0 fs, the present simulations were carried out in the same conditions as described above. The derivatives were evaluated at each of 12 fixed interatomic distances, corresponding to 0.5 Å intervals between 2.5 and 8.0 Å. One hundred picoseconds of MD sampling were employed at each window.

Because of the symmetric character of the present molecular system, it is anticipated that the free energy derivatives for the two carbon atoms will tend toward the same value after a certain period of simulation time. That this is not exactly the case, as may be observed in Figures 5 and 6, where the derivatives $\langle \partial H/\partial Q_i(\lambda) \rangle$ are plotted at an 8.0 and 2.5 Å intersolute separation, representing the end points of the simulation. It should be noted that to reduce the noise introduced by the equilibration stage, the averages were computed using the data in the reverse direction from that in which they were accumulated.

While the convergence is rapidly reached for $\langle \partial H/\partial R^*_{ij} \rangle$ and $\langle \partial H/\partial \varepsilon_{ij} \rangle$ (roughly 40 ps), a much larger amount of sampling seems to be required for $\langle \partial H/\partial q_i \rangle$. This is particularly true at the beginning of the PMF simulation (i.e., constraint distance of 8.0 Å), where, after 100 ps of sampling, the curves corresponding to the two carbon atoms have not yet converged toward a common value. At an intersolute distance of 2.5 Å, however, we can assume that convergence of $\langle \partial H/\partial q_i \rangle$ is attained after 60 ps. These facts may be correlated to the fluctuations induced by the motion of the water molecules between the "solvent-separated" solutes that generally causes the corresponding portion of the PMF profile to be noisier than the "contact" minimum area. This explains why it is so difficult to determine the depth of the "solvent-separated" minimum accurately.[19] This much slower convergence of $\langle \partial H/\partial q_i \rangle$, relative to $\langle \partial H/\partial R^*_{ij} \rangle$ and $\langle \partial H/\partial \varepsilon_{ij} \rangle$, is consistent with earlier observations for a system of two neon atoms in water.[19]

It has been demonstrated that one of the reasons for the possible poor convergence of free energy simulations is rooted in the lag between the Hamiltonian characterizing the current "λ" state and the actual configuration of the molecular system.[61] The consequence of this so-called Hamiltonian lag is that the accumulated data points at each window are not rigorously uncorrelated, although correlation between successive data points can exist regardless of such lag. Ideally, the variance of the mean of a series of N_{points} independent data points is expressed by:

$$\sigma^2(\bar{x}) = \frac{\sigma^2(x_i)}{N_{\text{points}}}; \tag{8}$$

in practice, this is almost never the case, and eq. (6) becomes[35]:

$$\sigma^2(\bar{x}) = (1 + 2\tau)\frac{\sigma^2(x_i)}{N_{\text{points}}} \tag{9}$$

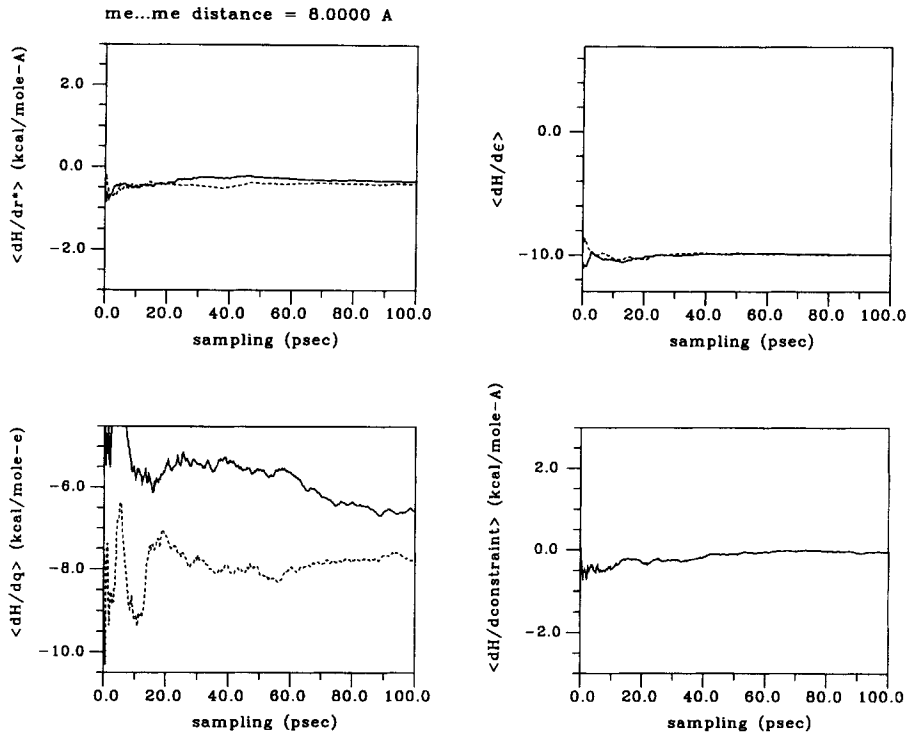

FIGURE 5. Free energy derivative averages as a function of the amount of MD sampling required for their evaluation in the case of two hydrated methane molecules separated by a carbon–carbon distance of 8.0 Å (i.e., beginning of the simulation). The derivatives for the two carbon atoms are depicted by solid and dashed lines, respectively.

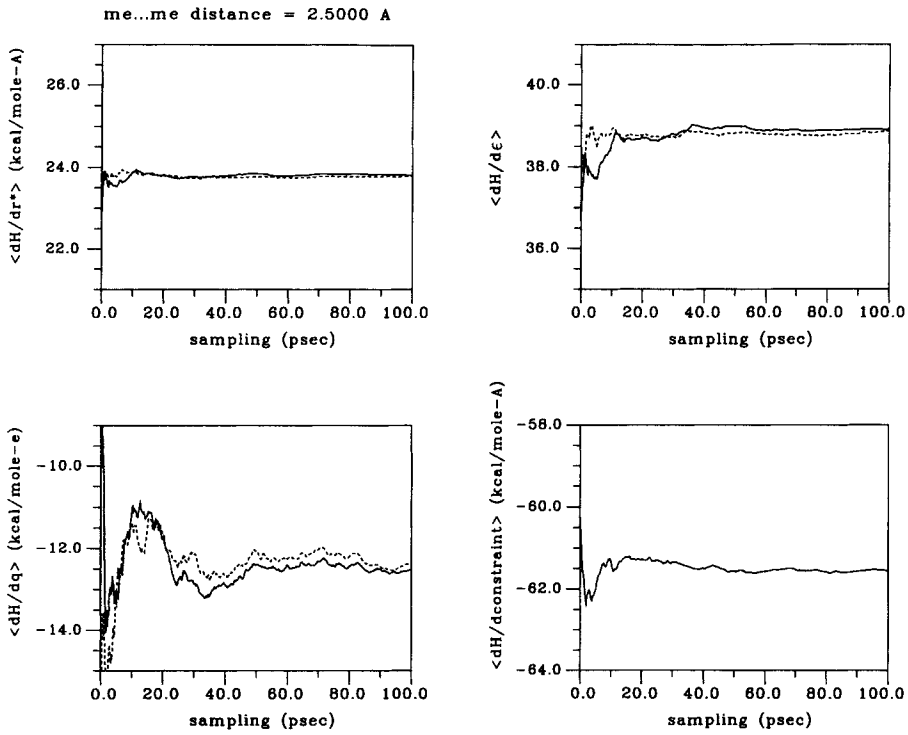

FIGURE 6. Free energy derivative averages as a function of the amount of MD sampling required for their evaluation in the case of two hydrated methane molecules separated by a carbon–carbon distance of 2.5 Å (i.e., end of the simulation). The derivatives for the two carbon atoms are depicted by solid and dashed lines, respectively.

where $(1 + 2\tau)$ is referred to as "sampling ratio."[62] Qualitatively, this quantity (although dimensionless) may be interpreted, when multiplied by δt, as the minimum amount of sampling that should be performed to assume that the collected data at individual windows correspond to truly independent observations. For each derivative, we have plotted the sampling ratio as a function of the constraint distance. While Pearlman[20] indicated a lower bound of the sampling time equal to 0.7 ps for systems involving the evaluation of the contribution due to constrained bonds, Figure 7 shows that the largest sampling ratio obtained for the methane–methane simulation is ca. 1.1 ps. This value is determined for $\langle \partial H/\partial r^{\text{constraint}} \rangle$ at a carbon–carbon distance of ca. 6.5 Å that is in the vicinity of the "solvent-separated" minimum. This remark is consistent with a prior observation suggesting that additional sampling is required in the areas where the entropic contribution due to the solvent is large.[19] In fact, we can easily assume—considering that the PMF profile starts at large intersolute separation (i.e., typically 8.0 Å)—that if not enough MD sampling is performed at each intermediate "λ" state, the error will propagate along the entire simulation. This assumption, among others, may explain the noise of the curves obtained over 1000 windows (0.5 ps of equilibration + 0.5 ps of data collection).

BENZENE–BENZENE PMF

The free energy profiles delineating the orthogonally constrained mutual approach of two benzene molecules are displayed in Figure 8a. These curves —derived from the three alternative methods FEP, TI, and SGTI using simulation protocols 16a, 16b, and 20—show analogous features. Qualitatively, each PMF possesses a deep and narrow global minimum located at a "contact" distance, together with a secondary shallower and wider minimum corresponding to a solvent separation of the two solutes. While the curves obtained from both the FEP and the TI simulations have a smooth appearance, the one derived using the more economical SGTI approach is conspicuously noisier. A detailed inspection of the different free energy profiles reveals that the three alternative methods provide consistent estimates of the depth of the solvent-separated minimum. As shown in Figure 8a this minimum occurs at ca. 8.2, 8.0, or 8.1 Å, with a respective free energy of −0.66, −0.74, or −0.85 kcal/mol, whether FEP, TI, or SGTI was employed. In contrast, the accord on the depth of the

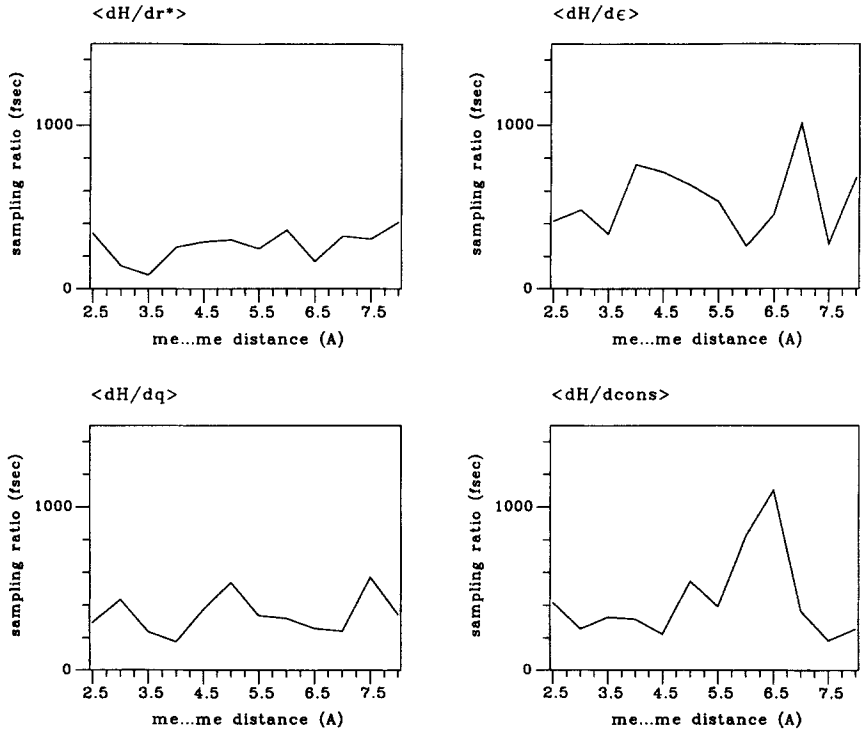

FIGURE 7. Sampling ratio[62] $(1 + 2\tau)\delta t$ as a function of the carbon–carbon interatomic distance R.

contact minimum is somehow less satisfactory as FEP, TI, and SGTI, respectively, predict an intersolute distance of ca. 4.9, 4.9, and 5.0 Å, with a corresponding free energy of −1.54, −1.94, and −1.69 kcal/mol. A comparison of the aforementioned three curves with the PMF derived by Linse[63] on a similar molecular system, using FEP, indicates that our "contact" distances are roughly 0.2 Å too large, while our solvent separations are 0.2 Å too short. On the other hand, the ca. −2.0 and −0.1 kcal/mol free energies estimated by Linse for the "contact" and the "solvent-separated" minima, respectively, are in good accord with the data provided by TI, although the potential energy function employed herein seems to lead to a stronger attraction at solvent separation. Does it mean that TI constitutes a better approach than FEP for this particular example?

While an unequivocal assessment of the convergence properties characterizing the present free energy simulations is relatively cumbersome, it is yet still possible to appreciate the accuracy of the derived profiles by examining the binding free energy of the T-shaped conformation at the contact distance. Just as for the hydrated methane dimer, we carried out FEP calculations to mutate $[(C_6H_6)_2]_{aqueous}$ into $[C_6H_6]_{aqueous}$, and $[C_6H_6]_{aqueous}$ into "nothing." The electrostatic decoupling procedure[2] was employed: 400 windows per leg, involving 0.5 ps of equilibration, followed by 0.5 ps of data collection per window. On account of the intrinsic problems that can arise when shrinking closed rings, the bond lengths of benzene were kept to their equilibrium values while the van der Waals parameters were annihilated. Employing this strategy, our estimate of the compositional change in free energy for the $[C_6H_6]_{aqueous}$ into nothing transformation is 0.40 ± 0.06 kcal/mol, which in view of the experimental value of 0.767 kcal/mol reported by Ben-Naim and Marcus,[60] falls nicely within the limits of chemical accuracy. On the other hand, we found that the mutation of $[(C_6H_6)_2]_{aqueous}$ into $[C_6H_6]_{aqueous}$ implies a free energy change equal to 2.10 ± 0.08 kcal/mol, so that the global binding free energy of benzene at "contact" distance (i.e., approximately 5.0 Å) amounts to −1.70 ± 0.14 kcal/mol. As can be seen in Figure 8a, this value agrees relatively well with the depths of the contact minima of the FEP, TI, and SGTI curves.

To further ascertain the quality and the accuracy of our results, we also evaluated the association constant K_a for each free energy profile, and compared it with the experimental value deter-

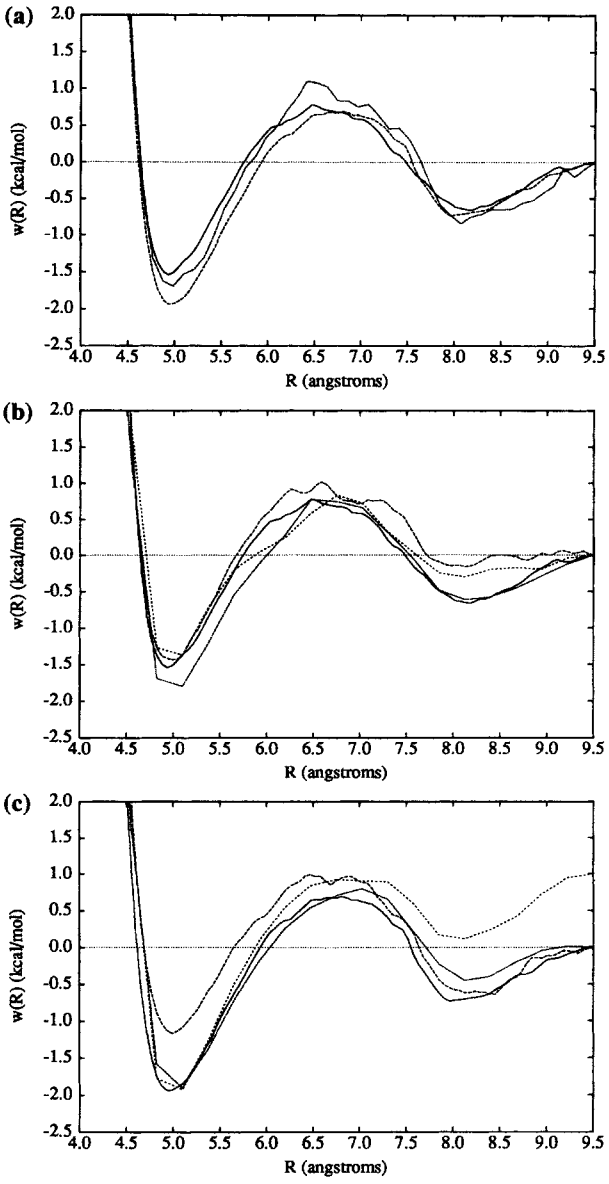

FIGURE 8. "T-shaped" benzene–benzene PMF evaluated using: (a) FEP (—) and TI (---), over 100 windows of 2.5-ps equilibration + 7.5-ps data collection, and SGTI, 100,000 windows of 2-fs data collection (···); (b) FEP and (c) TI over 100 windows of 2.5-ps equilibration + 7.5-ps data collection (—), 100 windows of 1.25-ps equilibration + 3.75-ps of data collection (– – –), 20 windows of 10-ps equilibration + 40-ps data collection (···), and 20 windows of 10-ps equilibration +15-ps data collection (---). All the profiles are generated as a function of R, the distance separating the noninteracting centroids of the aromatic rings.

mined earlier by Tucker and Christian.[64] The association constant for a dilute solution may be obtained by integrating[65,66] the PMF to an appropriate separation R_{cut}, such as:

$$K_a = 4\pi \int_0^{R_{cut}} r^2 e^{-w(r)/RT}\, dr. \qquad (10)$$

A second interesting key point of comparison between theory and experiment consists of estimating the osmotic second virial coefficient $\bar{B}$ from the MacMillan–Mayer theory:

$$\bar{B} = 2\pi \int_0^{\infty} r^2 [1 - e^{-w(r)/RT}]\, dr. \qquad (11)$$

We report in Table IV the series of K_a and $\bar{B}$, corresponding to the simulation protocols described previously. Compared to the experimental value[64] of 0.85 M^{-1}, it would seem that the value provided by FEP (i.e., roughly 1.42 M^{-1}) constitutes the best accord, and that derived from the TI curve (i.e., roughly 2.97 M^{-1}) the worst. From their studies of the orientationally averaged benzene–benzene PMF, Jorgensen and Severance,[67] followed by Linse,[68] estimated this association constant to be ca. 2 and 0.7 M^{-1}, respectively. As has been underlined by Jorgensen and Pranata, however, nonambiguous computation of association constants are more or less dependent on the cutoff value R_{cut} appearing in eq. (10). In this study, we set this parameter to ca. 6.0 Å, which corresponds approximately to the limit of the "contact" minimum. Conversely, the formulation of the osmotic second virial coefficient and its connection with the free energy profile leaves no ambiguity in the integral boundaries. As shown in Table IV, the best agreement with the experimental values of -1177 Å^3 and -1001 Å^3 $< \bar{B} <$ -276 Å^3, reported earlier by Tucker and Christian[64] and Rossky and Friedman,[69] respectively, is given by FEP (ca. -1430 Å^3). In contrast, the PMF curve derived from the TI simulation, and, to a lesser extent from the SGTI one, leads to the worst accord with -2835 Å^3.

TABLE IV.
Association Constant K_a and Osmotic Second Virial Coefficient $\bar{B}$ of Benzene in TIP3P Water.

Simulation[a]	K_a (M^{-1})	$\bar{B}$ (Å^3)
16a	1.42	−1430
16b	2.97	−2835
17a	1.20	−412
17b	0.78	−761
18a	3.01	−2695
18b	3.35	−2433
19a	1.43	−866
19b	3.42	−1622
20	2.04	−2295

[a] See simulation protocols in Table II.

In an attempt to probe the effect of the window width $\delta\lambda$ on the quality of the free energy profiles, we considered a series of MD simulations involving a very limited number of intermediate states, namely 20. The set of curves reproduced in Figure 8b were derived using sampling protocols 16a, 17a, 18a, and 19a. From the onset, the four free energy profiles show qualitatively analogous features. In particular, they all possess two minima of similar respective depths. A closer look reveals that the agreement between the "solvent-separated" minima obtained from simulations 16a and 18a is excellent. In consideration of the extensive sampling performed at each of the 20 windows, this fact is, however, not totally surprising. Conversely, assigning the *exact* position and depth of the contact minimum of the curves generated with few intermediate "λ" states may, under certain circumstances, be quite problematic. The constrained T-shaped benzene dimer constitutes a good illustration, as its "contact" minimum turns out to be relatively deep and narrow, and thus requires an appreciably larger number of windows to reproduce its shape faithfully. Nevertheless, the "contact" minima of the PMFs derived over 500 ps and 1 ns (cf. simulation strategies 18a and 19a) appear to bracket with approximately ± 0.3 kcal/mol the minimum of the profile characterizing simulation 16a. The curve constructed from simulation 17a is certainly noisier than the "reference" free energy profile obtained also over 100 windows, but twice as much sampling between each interval was performed. Quantitatively, the poor resolution of the PMF in its "solvent-separated" area makes the estimation of the minimum depth relatively difficult. Nevertheless, it should be pointed out that the quality is far better at "contact" distance, and that the derived association constant and osmotic second virial coefficient are in good accord with the experiment (see Table IV).

The ensemble of free energy profiles depicted in Figure 8c were obtained from simulations 16b, 17b, 18b, and 19b. The most striking feature emerging from this set of curves is the seemingly

diverging PMF that was generated over a total simulation time of 500 ps (cf. protocol 20b). In fact, we adjusted the profile vertically so that its "contact" matches those obtained over longer simulation times. What this curve suggests is a propagation of the error along the reaction path, as an obvious result of insufficient sampling at each individual "λ" point of the integrand. Compared to the FEP profile derived employing an identical simulation strategy, it would seem that TI is somewhat more sensitive, not only to the λ spacing, but also, above all, to the amount of sampling performed at each window. In view of the results described in the preceding sections, this may not be necessarily true, and the *a priori* reasonable free energy profile obtained from the short FEP simulation 19a may be purely artifactual. Another point of interest is provided by the PMF derived using simulation strategy 17b. Although noisier than the profile corresponding to an identical $\delta\lambda$ but involving twice as much MD sampling, the accord for the solvent-separated minimum is still fairly acceptable. In contrast, both the height of the barrier separating the two minima and the depth of the contact minimum seem erroneous. It is likely, in our view, that the underestimation by ca. 0.9 kcal/mol of the depth of the minimum occurring at contact distance stems from inadequate sampling.

From the data presented in this section, it appears again that FEP and TI lead to equivalent results, under the *sine qua non* condition that enough MD sampling be performed at each individual intermediate "λ" state. This is well illustrated by the free energy profiles derived from simulations 16a, 16b, 18a, and 18b. From a purely quantitative standpoint, discrepancies in the estimation of the "contact" minimum depth still exist, and suggest that convergence may not have been attained completely yet. On the other hand, the smooth appearance of the reproduced curves, combined to the admittedly good agreement between the derived association constant K_a and osmotic second virial coefficient $\overline{B}$ and the experiment,[64,69] tend to indicate that the level of accuracy of our simulations is not incompatible with a quantitative interpretation of the results. It should nonetheless be clearly emphasized that unequivocally converged free energy calculations would require even longer total simulation times, which, considering the CPU times reported in Table III, could be extremely prohibitive. In particular, unlike simple PMF simulations requiring a single holonomic distance constraint to bring the two solutes together, argon–argon or methane–methane, the present computations involve additional angular constraints insuring that the dimer will remain in its T-shaped conformation along the entire reaction path. Because of the nonsymmetric character of these constraints, a substantial amount of the total CPU time will be spent in the dedicated subroutines so that all the constraints are satisfied simultaneously. While FEP and TI imply similar timings in such subroutines, the actual generation of the PMF turns out to be significantly faster in the case of TI—a fact that could influence the choice of one computational approach over the other.

Conclusion

In this study we investigated alternative approaches to determine the PMF of three distinct systems, of increased complexity, namely: two tagged argon atoms in liquid argon; two methane molecules; and two benzene molecules, both in an aqueous solution. The ensemble of MD simulations reported herein, using a variety of sampling strategies, tend to prove that FEP and TI yield globally analogous results. However, comparing curves generated using the same number of $\delta\lambda$ intervals and the same amount of sampling at each interval, the FEP method often appears to be slightly poorer than TI for the generation of PMF profiles. In particular, the relatively noisier curves frequently observed from *a priori* long-enough FEP simulations reflect *not only* the somewhat greater sensitivity of the method to the orthogonality of the Hamiltonians characterizing contiguous "λ" states, *but also* the necessity to perform a larger amount of MD sampling at each given window. Interestingly enough, considering different molecular systems, Van Gunsteren[70] and subsequently Straatsma and McCammon[15] formed similar postulates regarding the superiority of the alternative TI approach over the hitherto widely utilized FEP method.

FEP and TI constitute robust computational approaches that can give reliable and accurate results, employing a relatively limited number of intermediate "λ" states (typically 50 windows), provided that the MD sampling per window is sufficient. Free energy profiles of acceptable quality can also be obtained by reducing the window width $\delta\lambda$ substantially, so that adjoining "λ" states were very similar. Sampling protocols between these two extremes correspond to situations where

the combination of the number of windows with the amount of sampling associated to them clearly plays a crucial role on the overall convergence characteristics of the molecular simulations. The PMF curves presented here, using such intermediate strategies, would seem to indicate that TI is slightly safer than FEP; but, considering that none of the simulations appear to have converged unequivocally, it is legitimate to wonder whether this issue is really critical. In fact, what really emerges from this survey is that proper convergence of the free energy calculations and quantitative interpretation of the results require *extremely* long simulations. We have demonstrated unambiguously that, if such is the case, FEP and TI yield very comparable free energy profiles, the key factor that could motivate the choice of one method over the other being essentially related to computational efficiency. In this respect, we have shown that TI consistently requires slightly smaller CPU investments than FEP.

Both the comparison of the FEP and TI free energy profiles with the "exact" PMF, when available, and the analysis of the free energy derivatives with respect to the atomic parameters along the reaction path, confirm that extensive sampling is required to reach appropriate convergence. In agreement with prior studies,[19] the comparison of the PMF curves derived from TI simulations—and, to a certain extent, FEP—using different sampling protocols tends to indicate that it is somehow preferable to employ a reduced number of windows involving a large amount of sampling, rather than numerous windows with fewer samplings. In particular, an inspection of the sampling ratios evaluated at several intersolute distances supports the view that convergence is more difficult to reach in the area corresponding to the solvent-separated minimum. For this portion of the PMF profile, it is recommended to increase the amount of sampling significantly to ensure an accurate estimate of the minimum depth.

Finally, in the context of PMF simulation of small solutes, the so-called SGTI approach seems to yield promising results at a much lower cost than either conventional TI or FEP calculations. However, the assumption that the requisite ensemble average at any window can be approximated by a very limited number of instantaneous values, calls into question the reliability of the results. In our view, any finite alteration of the classical Hamiltonian—even markedly small—will perturb the postulated continuous equilibrium within the periodic box; such equilibrium can only be recovered after appropriate sampling to allow complete relaxation of the system. While this aspect is anticipated to play a minor role in the case of nonpolar solutes for which the interactions with the solvent are generally weak, it is not clear yet whether the alternative SGTI method can be applied safely to all molecular systems.

Acknowledgments

The authors wish to thank Dr. C. Millot for many stimulating discussions, and the continuous interest and encouragement of Professor B. Maigret. We also thank Dr. James C. Caldwell for providing his customized version of EDIT / AMBER 4.0.[71] C. C. is indebted to Roussel Uclaf Company (France) for his doctoral fellowship, and P. A. K. is grateful to the NSF (CHE-91-13472) for research support.

References

1. J. P. M. Postma, H. J. C. Berendsen, and J. R. Haak, *Faraday Symp. Chem. Soc.*, **17,** 55 (1982).
2. P. A. Bash, U. C. Singh, R. Langridge, and P. A. Kollman, *Science*, **236,** 564 (1987).
3. D. L. Beveridge and F. M. DiCapua, *Annu. Rev. Biophys. Biophys. Chem.*, **18,** 431 (1989).
4. W. L. Jorgensen, In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, W. F. Van Gunsteren and P. K. Weiner, Eds., Escom, The Netherlands, 1989, p. 60.
5. W. L. Jorgensen, *Acc. Chem. Res.*, **22,** 184 (1989).
6. D. L. Beveridge and F. M. DiCapua, In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, W. F. Van Gunsteren and P. K. Weiner, Eds., Escom, The Netherlands, 1989, p. 1.
7. D. A. Pearlman and P. A. Kollman, In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, W. F. Van Gunsteren and P. K. Weiner, Eds., Escom, The Netherlands, 1989, p. 101.
8. W. F. Van Gunsteren and H. J. C. Berendsen, *Angew. Chem. Int. Ed. Engl.*, **29,** 992 (1990).
9. R. H. Wood, *J. Phys. Chem.*, **95,** 4838 (1991).
10. J. A. McCammon, *Curr. Opin. Struct. Biol.*, **1,** 196 (1991).
11. A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, Wiley–Interscience, New York, 1991.
12. T. P. Straatsma and J. A. McCammon, *Annu. Rev. Phys. Chem.*, **43,** 407 (1992).
13. P. A. Kollman, *Chem. Rev.*, **93,** 2395 (1993).
14. D. A. Pearlman and P. A. Kollman, *J. Chem. Phys.*, **94,** 4532 (1991).
15. T. P. Straatsma and J. A. McCammon, *J. Chem. Phys.*, **95,** 1175 (1991).

16. M. J. Mitchell and J. A. McCammon, *J. Comput. Chem.*, **12,** 271 (1991).
17. M. Mazor and B. M. Pettitt, *Mol. Simul.*, **6,** 1 (1991).
18. S. H. Fleischman and D. A. Zichi, *J. Chem. Phys.*, **88,** 2617 (1991).
19. D. A. Pearlman, *J. Chem. Phys.*, **98,** 8946 (1993).
20. D. A. Pearlman, *J. Comput. Chem.*, **15,** 105 (1994).
21. D. A. Pearlman, *J. Phys. Chem.*, **98,** 1487 (1994).
22. C. Chipot, C. Millot, B. Maigret, and P. A. Kollman, *J. Phys. Chem.*, **98,** 11362 (1994).
23. M. Karplus and J. A. McCammon, *Annu. Rev. Biochem.*, **52,** 263 (1983).
24. W. L. Jorgensen, *J. Phys. Chem.*, **87,** 5304 (1983).
25. R. W. Zwanzig, *J. Chem. Phys.*, **22,** 1420 (1954).
26. M. Mezei and D. L. Beveridge, *Ann. NY Acad. Sci. USA*, **482,** 1 (1986).
27. M. R. Mruzik, F. F. Abraham, D. E. Schreiber, and G. M. Pound, *J. Chem. Phys.*, **64,** 481 (1976).
28. M. Mezei, S. Swaminathan, and D. L. Beveridge, *J. Am. Chem. Soc.*, **100,** 3255 (1978).
29. R. H. Wood, W. C. F. Muhlbauer, and P. T. Thompson, *J. Phys. Chem.*, **95,** 6670 (1991).
30. T. Simonson and A. T. Brünger, *Biochemistry*, **31,** 8661 (1992).
31. H. A. Carlson, T. B. Nguyen, M. Orozco, and W. L. Jorgensen, *J. Comput. Chem.*, **14,** 1240 (1993).
32. C. Chipot, C. Millot, B. Maigret, and P. A. Kollman, *J. Chem. Phys.*, **101,** 7953 (1994).
33. C. H. Bennett, *J. Comp. Phys.*, **22,** 245 (1976).
34. W. L. Jorgensen and C. Ravimohan, *J. Chem. Phys.*, **83,** 3050 (1985).
35. T. P. Straatsma, H. J. C. Berendsen, and A. J. Stam, *Mol. Phys.*, **57,** 89 (1986).
36. D. J. Tobias and C. L. Brooks III, *Chem. Phys. Lett.*, **142,** 472 (1987).
37. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.*, **79,** 926 (1983).
38. D. A. Pearlman, D. A. Case, J. C. Caldwell, W. S. Ross, T. C. Cheatham, G. Seibel, U. C. Singh, P. Weiner, and P. A. Kollman, *AMBER 4.1*, University of California, San Francisco (UCSF), San Francisco, 1994. The new version of GIBBS employed in this work includes, among others, the possibility of carrying out PMF calculations using the TI method.
39. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta Jr., and P. Weiner, *J. Am. Chem. Soc.*, **106,** 765 (1984).
40. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comput. Chem.*, **7,** 230 (1986).
41. W. D. Cornell, P. Cieplak, C. I. Bayly, and P. A. Kollman, *J. Am. Chem. Soc.*, **115,** 9620 (1993).
42. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. C. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.*, **117,** 5179 (1995).
43. A. T. Hagler, E. Euler, and S. Lifson, *J. Am. Chem. Soc.*, **96,** 5319 (1974).
44. S. R. Cox and D. E. Williams, *J. Comput. Chem.*, **2,** 304 (1981).
45. C. Chipot and J. G. Ángyán, *GRID Version 3.0: Point Multipoles Derived from Molecular Electrostatic Properties*, QCPE No. 655, Indiana University, 1994.
46. P. C. Hariharan and J. A. Pople, *Chem. Phys. Lett.*, **16,** 217 (1972).
47. H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. DiNola, and J. R. Haak, *J. Chem. Phys.*, **81,** 3684 (1984).
48. H. C. Andersen, *J. Chem. Phys.*, **72,** 2384 (1980).
49. J. Ryckaert, G. Cicotti, and H. J. C. Berendsen, *J. Comput. Phys.*, **23,** 327 (1977).
50. W. F. Van Gunsteren and H. J. C. Berendsen, *Mol. Phys.*, **34,** 1311 (1977).
51. D. J. Tobias and C. L. Brooks III, *J. Chem. Phys.*, **89,** 5115 (1988).
52. C. Pangali and B. J. Berne, *J. Chem. Phys.*, **71,** 2975 (1979).
53. G. Ravishanker, M. Mezei, and D. L. Beveridge, *Faraday Symp. Chem. Soc.*, **17,** 79 (1982).
54. K. Watanabe and H. C. Andersen, *J. Phys. Chem.*, **90,** 795 (1986).
55. W. L. Jorgensen, J. K. Buckner, S. Boudon, and J. Tirado-Rives, *J. Chem. Phys.*, **89,** 3742 (1988).
56. C. X. Wang, H. Y. Liu, Y. Y. Shi, and F. H. Huang, *Chem. Phys. Lett.*, **179,** 475 (1991).
57. L. R. Pratt and D. Chandler, *J. Chem. Phys.*, **67,** 3683 (1977).
58. L. R. Pratt and D. Chandler, *J. Chem. Phys.*, **73,** 3430 (1980).
59. L. R. Pratt and D. Chandler, *J. Chem. Phys.*, **73,** 3434 (1980).
60. A. Ben-Naim and Y. Marcus, *J. Chem. Phys.*, **81,** 2016 (1984).
61. D. A. Pearlman and P. A. Kollman, *J. Chem. Phys.*, **91,** 7831 (1989).
62. E. B. Smith and B. H. Wells, *Mol. Phys.*, **52,** 701 (1984).
63. P. Linse, *J. Am. Chem. Soc.*, **114,** 4366 (1992).
64. E. E. Tucker and S. D. Christian, *J. Phys. Chem.*, **83,** 246 (1979).
65. J. E. Prue, *J. Chem. Ed.*, **46,** 12 (1969).
66. D. Shoup and A. Szabo, *Biophys. J.*, **40,** 33 (1982).
67. W. L. Jorgensen and D. L. Severance, *J. Am. Chem. Soc.*, **112,** 4768 (1990).
68. P. Linse, *J. Am. Chem. Soc.*, **115,** 8793 (1993).
69. P. J. Rossky and H. L. Friedman, *J. Phys. Chem.*, **84,** 587 (1980).
70. W. F. Van Gunsteren, In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, W. F. Van Gusteren and P. K. Weiner, Eds., Escom, The Netherlands, 1989, p. 27.
71. D. A. Pearlman, D. A. Case, J. C. Caldwell, G. Seibel, U. C. Singh, P. Weiner, and P. A. Kollman, *AMBER 4.0*, University of California, San Francisco (UCSF), San Francisco, 1991.